The New *Geology Ontario* Web Portal— An "out-of-the-box" Solution for Discovering and Delivering Ontario's Geoscience Data

By Zoran Madon[1], P.Geo. and Brian Berdusco[2], M.Sc.

[1]Ontario Geological Survey (OGS)/Ministry of Northern Development and Mines
Willet Green Miller Centre, B7
933 Ramsey Lake Road
Sudbury, ON P3E 6B5
Telephone: (705) 670-5991
Fax: (705) 670-5905
e-mail: zoran.madon@ontario.ca

[2]Business Solutions/Land and Resources Cluster
Ministry of Northern Development and Mines (MNDM)
159 Cedar Street, Suite 605
Sudbury, ON P3E 6A5
Telephone: (705) 564-7093
Fax: (705) 564-7919
e-mail: Brian.Berdusco@ontario.ca

INTRODUCTION

The new *Geology Ontario* web portal (www.ontario.ca/geology) was developed during the last two years as a direct result of the Minister's Office directive to improve the discovery and delivery of Ministry of Northern Development and Mines (MNDM) geoscience data (Figure 1). This directive came about in response to clients' complaints about MNDM's existing delivery mechanisms. Many clients work in mining jurisdictions throughout North America and around the world, and are accessing geoscience data in a variety of ways. Not surprisingly, they quickly adopt the best methods for accessing data and, subsequently, demand this high level of service delivery from other jurisdictions.

BACKGROUND

The GeoPortal Project was initiated in the spring of 2004. The project team, made up of a project manager and a small group of individuals seconded from various sections within MNDM, held several meetings initially to scope out the requirements and to assign responsibilities.

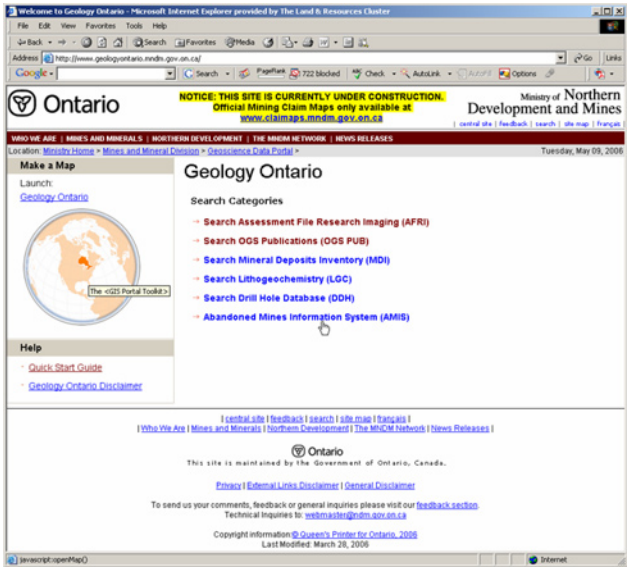


Figure 1. ***Geology Ontario*** home page allowing access to both the map search window as well as the text search windows for the MNDM's geoscience data archives (AFRI, PUBS, MDI, LGC, ODHDB and AMIS).

The project specifications were constrained by the following core requirements:

- create a delivery website using an "out-of-the-box" solution that meets current needs
- concentrate on delivering all OGS publications (PUBS), mineral exploration assessment files (AFRI), the Mineral Deposit Inventory (MDI), the Ontario Drill Hole Database (ODHDB), the Lithogeochemical Database (LGC), land tenure information (*CLAIMaps III*), and Abandoned Mines Information (AMIS) – totaling in excess of 300 gigabytes of data including 2 million text pages and 168,000 maps.

Work on the *Geology Ontario* web portal began on several fronts, including the selection of a software vendor to provide the web service, a benchmarking study, a client survey, development of appropriate metadata records for all data holdings, and conversion of various existing digital image files to a more useful, popular and web-friendly format. This initial work was followed up with business-impact and threat-risk studies, focus group sessions, and a Quality Assurance/Quality Control process to test the web portal functionality.

Software Vendor Selection Process

In 1997, MNDM consolidated nine regional Provincial Mining Recorder Offices into a single, centralized office in Sudbury, Ontario. This necessitated a new way to distribute 40,000 mining land tenure (claim) maps annually, as well as a new process for maintaining information on these maps. The Ministry decided to use the internet as a service delivery mechanism, and as a result, the Crown Land Automated Internet Mapping System (*CLAIMaps*) was developed. *CLAIMaps*, originally a simple ArcView IMS application, delivered scanned images of claim maps over the internet. However, the process of updating these maps was limited to editing hard copies, rescanning them, and then reposting them on the Ministry's website.

In 2000, the decision was made to update the CLAIMaps application based on Environmental Systems Research Institute (ESRI) Map Objects Internet Map Service (MOIMS) software. The new application, CLAIMaps II, quickly grew in use, and the heightened demand placed on the system resulted in critical, technical challenges on server infrastructure. In addition, the solution was based on complex custom code that required significant human and financial resources to maintain. This system ran from March 2001 to December 2002.

The next phase of the application was built on an out-of-the-box philosophy, and a new *CLAIMaps* website based on ESRI ArcSDE, ArcIMS, and Oracle RDBMS was launched. The new system, *CLAIMaps III*, turned out to be a stable application that required few resources to maintain and resulted in a greatly reduced total cost of ownership. It was also recognized internationally as a world class system capable of generating and delivering daily-updated map images and polygon data (Figure 2).

The success of CLAIMaps III made ArcSDE, ArcIMS, and Oracle the obvious choices as the base for building *Geology Ontario*. The Ministry approached companies on the provincial Vendor of Record and requested a time and materials quote for developing the new portal. Companies had to show a proficiency in developing web-based applications based on ESRI ArcSDE, ArcIMS, Oracle RDBMS, and the ArcIMS Portal Toolkit. Like *CLAIMaps*, the development environment for *Geology Ontario* was the Java 2 Plaform, Enterprise Edition (J2EE).

Benchmarking Study

Benchmarking is a crucial step in helping define customer requirements, establishing objectives, developing productivity measures, and remaining competitive. In this benchmark study, over a hundred geoscience web sites were examined in order to document industry best practices and ensure that the *Geology Ontario* web portal would provide the necessary services to meet client requirements. Data discovery methods (i.e., map search tools, text search tools), data content, and data format were tabulated for each site visited.

Over 50% of the sites offer text-based search tools, while more than 30% offer map-based search tools. Approximately 25% of these use ArcIMS technology. The rest maintain either simple listings in some order (i.e., chronological, alphabetical, by subject matter) because of the low volume of downloadable data or do not offer any data for download, but simply provide a link to their publications office. Most of the sites with downloadable data have some form of metadata, and many of these are FGDC/Z39.50 compliant. Some sites even produce thumbnail images of the document or map.

Almost all of the downloadable sites supply geoscience publications, both maps and reports, as well as regional thematic data that encompasses their entire jurisdiction. These would typically include regional planimetric, geological and geophysical maps, mineral deposit and regional geochemical data, and in some cases regional terrain models (DEMs) and satellite imagery. Some sites offer mineral land tenure maps and assessment file reports, as well.

By far, the most popular format for disseminating both reports and maps is the Adobe Systems PDF (Portable Document Format). GIS and vector data downloads are available in various formats, but the ESRI Arcview/ArcInfo formats predominate. Tabular data stored in spreadsheets is generally provided in Microsoft Excel and ASCII CSV formats. Very few sites supply image and grid data

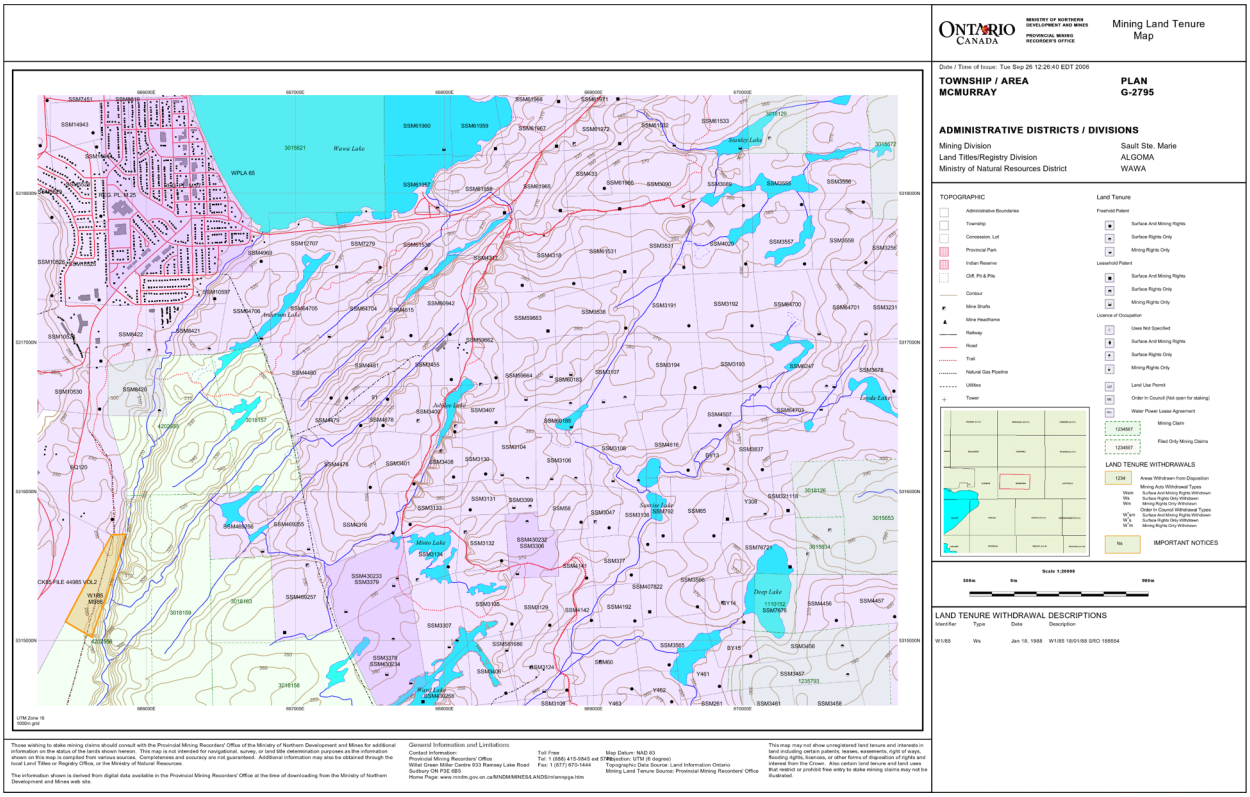


Figure 2. Sample of a typical map product generated from the *CLAIMaps III* web site.

(satellite images, geophysical data), so no dominant format emerged here. Rather, they appear to be spread evenly across the following types – Lizardtech MrSID, Geosoft GRD/GXF, ERMapper ERS, TIFF, JPG, and BMP.

All sites, with the exception of one, provide free downloads. Some sites require individuals to register on-line before being allowed to download, particularly sites that provide access to very large files such as geophysical data and satellite imagery. Disclaimers and liability waivers vary from the non-existent to those that appear before each download. Some sites offer the possibility of downloading data in more than one format and, for maps, in more than one projection. Most sites provide only current data for download – generally less than 5 to 10 years old.

Client Survey

A client survey was posted on MNDM's e-consultation web site for approximately 6 weeks in the summer of 2004. The survey consisted of 26 questions, primarily in multiple choice format, and generally took less than 10 minutes to complete. Throughout the questionnaire, respondents were given the opportunity to comment, critique, and provide suggestions. The survey was subdivided into the following sections:

- background information (business activity, geographic area of interest, internet connection used, etc.),
- data discovery/access preferences (map-based vs. text-based searches, text-based search criteria),
- data content/format preferences (which government geoscience data sets were most important, what formats do you use for report data, map data, GIS data, tabular data, etc.)
- data usage (plotting on maps, incorporating into reports, spatial modeling, reference material, etc.) and
- data quality issues (method for reporting data errors, etc.).

Approximately 200 respondents completed the survey, which provided the project team with a better profile of clients' needs and preferences as well as a wealth of constructive suggestions to incorporate into the web design. Suggestions included the following requests: list examples or provide pick lists for all text-search criteria, develop an on-line help utility, produce hyperlinks to documents directly from the search results page, and provide an easy link to communicate problems or issues pertaining to the web portal.

Metadata Development and Data Conversion Process

Metadata pertaining to MNDM publications is currently housed in two different databases. In 1994, the Ministry developed a database that contains attributes on all publications produced since 1891. In 2002, the Ministry began entering metadata into a centralized metadata server based on the Government of Ontario Information Technology Standard (GOITS) 72.0. GOITS 72.0 is a metadata standard that was derived from the U.S. Federal Geographic Data Committee (FGDC) standards. Though new efforts have concentrated on the centralized metadata server, MNDM continues to populate the older publications database simply because this database contains records for all publications.

The Ministry has created a process of extracting metadata from the publications database and populating the metadata catalog available in ArcIMS. In this implementation, MNDM has utilized the FGDC compliant catalog but will review moving to the new ISO standard in the near future. The metadata server is a Z39.50 protocol compliant server that enables direct access from library search engines.

Metadata values were also used to populate fields within the PDF documents themselves. The purpose of populating these fields or meta-tags is to facilitate the rapid discovery of pertinent information either through the website search engine or through other common search engines such as Google and Yahoo. These search engines not only index content, but also index information contained within PDF meta-tags. The Ministry also embedded meta-tag data within the mineral assessment files to facilitate discovery.

The majority of publications and assessment files were scanned in the early 1990s and stored as raster images. MNDM utilized raster to PDF conversion software available from JRAPublish and converted approximately 2,000,000 pages and 168,000 maps to 85,000 PDF documents. These documents, in turn, were subjected to optical character recognition software and converted to "PDF searchable" format. The Ministry built a 40,000 entry geoscientific and geographic names dictionary to enhance word pattern recognition by the conversion software. The net result is that the majority of these millions of raster images can now be searched for content using the dtSearch engine that has been implemented on the *Geology Ontario* website.

Simple and complex Boolean functions coupled with meta-tag search fields ensure that clients are provided with simple tools that return powerful results (Figure 3). For example, using the following Boolean syntax: *visible w/3 gold* returns those occurrences where the word *visible* is found within three words of the word *gold*. Though this may seem like an obvious query to perform, what must be realized is that performing this type of query will also return occurrences of *no visible gold*. A simple modification of the above query to: *visible w/3 gold not w/2 no* will remove the occurrences of *no visible gold*. Combining these Boolean queries with meta-tag data can assist the client in narrowing down results and pinpointing searches.

In the future, metadata will probably be entered, stored, and managed through ESRI ArcCatalog. Not only will searches align with the Z39.50 protocol, but the data distributed through the website will also include necessary attributes for proper spatial queries. ArcCatalog also conforms to the out-of-the-box philosophy that has contributed to a reduction in the total cost of the development and maintenance of internet-facing, GIS-related web applications.

Business-Impact Analysis and Threat-Risk Analysis

As work on *Geology Ontario* progressed, the project team undertook a business-impact analysis, a threat-risk analysis, and a business continuity plan, using guidelines developed by the Ontario government's business audit group.

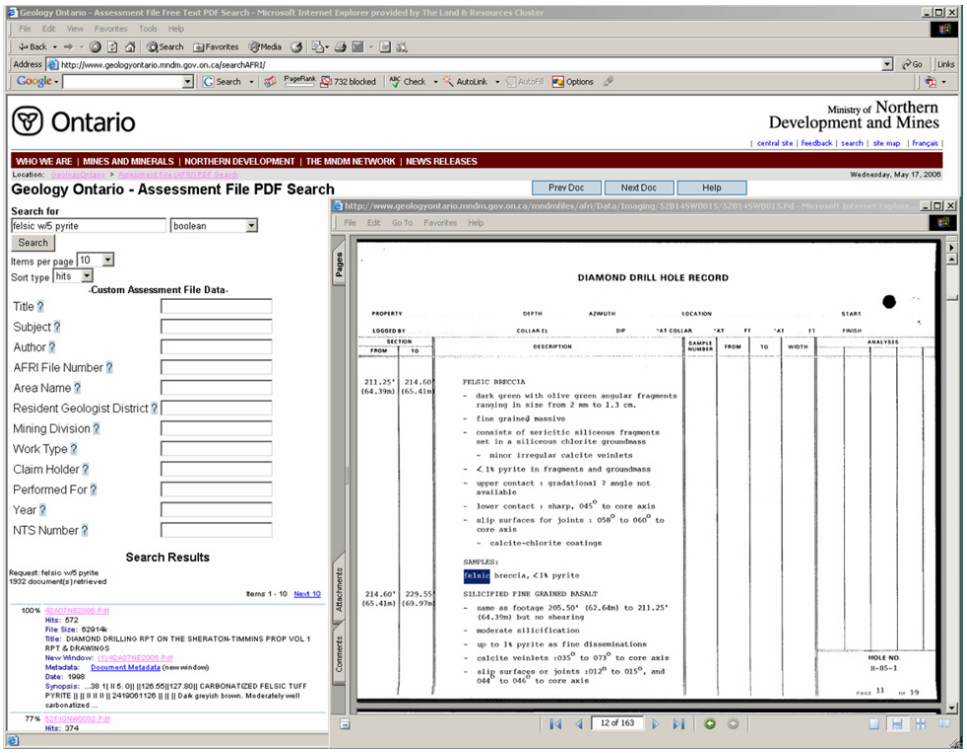
The business impact study attempts to evaluate how an interruption in the *Geology Ontario* web service would impact clients over time. Although several client impact variables are available in the guidelines (including public and employee health and safety, business service delivery, public confidence, internal perception, regulatory and legal, environmental impact, financial reporting, revenue loss), it was concluded that disruption of the web service would have minimal impact for only business service delivery in the short term (< 1 day) and moderate to high impact for public confidence and internal perception in the longer term (2 weeks to 1 month).

The threat-risk analysis evaluated all the *Geology Ontario* assets, including data, software/hardware architecture, network, and staff. For each asset, a sensitivity assessment with respect to confidentiality, integrity, availability, and accountability was estimated to determine the potential severity of harm if the asset was compromised in any way. Threat agents, both internal (i.e., users, system administrators) and external (i.e., clients, hackers, contractors, natural and man-made disasters, etc.), as well as threat events, were identified. Once determined, an estimate for the likelihood (i.e., low, moderate or high) of each threat event as well as its exposure (i.e., the impact to the government if the threat is realized) was established.

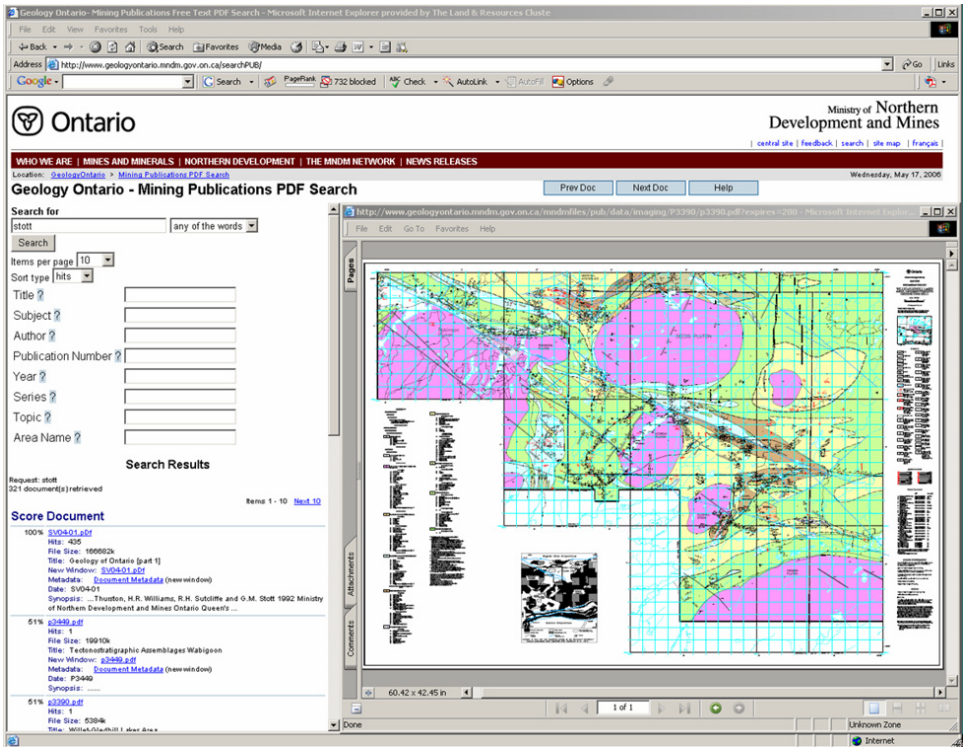
Finally, a business continuity plan was developed to recommend alternate strategies for delivering *Geology Ontario* services during an unexpected interruption.

Focus Group

When a beta version of the website was available for review and testing, three individuals who represented



A



B

Figure 3. ***Geology Ontario*** "PDF searchable" files allow for advanced search capability; A) in this example, a Boolean search for the word "felsic" within 5 words of "pyrite" returned almost 2000 hits from the ***Geology Ontario*** AFRI archives; B) in this example, a simple search using an OGS geologist's last name returned over 300 documents, including maps from the ***Geology Ontario*** PUBS archive.

MNDM's main client groups were invited to Sudbury for a "test drive." After a few sessions, the individuals completed a Ministry evaluation form to provide feedback on all aspects of the *Geology Ontario* web portal, including:

- the portal home page functionality, layout, and content
- on-line help
- layout and functionality of the search pages for the various themes
- layout and functionality of the map viewer window
- overall search functionality
- layout and functionality of the search results window
- functionality of the download process
- general comments when navigating between different windows.

Their input helped to further refine and improve the ***Geology Ontario*** web portal.

Quality Assurance and Quality Control

The *Geology Ontario* website replaces the Earth Resources Mineral Exploration webSite (ERMES). However, since both *Geology Ontario* and ERMES operate on copies of the same Oracle databases, quality assurance processes were developed whereby query results generated in ERMES were compared with the same results generated in *Geology Ontario*. Any discrepancies between the two pointed to possible inherent problems within query structures. In some cases, it was discovered that the original queries in ERMES were incorrect.

Data integrity issues were a persistent problem in both the OGS Publications and AFRI databases. Historically, files were loaded using a complex data loading process developed in the early 1990s, which involved populating Oracle databases and storing raster images of individual pages and maps. It was not uncommon for data loading problems to cause orphaned attribute or image data. Orphaned documents could not be queried and would remain, essentially, undiscoverable. However, once these scanned images were converted to searchable PDF format and indexed, they became discoverable through the text search engine. Orphaned files have now been identified and targeted for correction so that both attribute and text-based queries return similar results.

The original scanned documents from the early 1990s were stored in either JPG or TIFF format, and as thumbnail GIF images. On occasion, the original TIFF images were scanned and stored incorrectly, which rendered them unconvertible to PDF. MNDM worked with the PDF software vendor (James Rile Associates) and developed a process whereby a specially designed page was inserted in place of any unusable TIFF image. Embedded within the "bad page" notice was a special code that was, in turn, indexed by the text search engine. By searching for this code, a record was generated that indicated which file pages were troublesome. The "bad page" notice, therefore, serves the dual purpose of informing clients that a page was scanned incorrectly and failed to convert, and assisting the development and maintenance team in identifying which pages required re-scanning.

JPG images that were created at that time failed to populate the data pertaining to resolution (dots per inch or dpi) in the JPG header. The PDF conversion application defaulted to an incorrect dpi which, when the JPGs were converted, caused them to display at the wrong scale. Since the original scanning was done at 200 dpi for JPG images, an image conversion utility was run to update the headers in every JPG file with the 200 dpi setting. As a result, these images converted to the correct scale in the final PDF document.

The Ministry will develop a problem report page where clients can submit information on any problems they encounter with products downloaded from the ***Geology Ontario*** website. In addition, clients can recommend changes that would assist them in using the web-based products.

GEOLOGY ONTARIO WEB PORTAL

Benefits of Out-of-the-Box Software

The entire application development environment adhered to the same "out-of-the-box" philosophy that has made the *CLAIMaps* website so successful. Minimal custom code generation will facilitate the Ministry's ability to manage and modify the *Geology Ontario* website on an ongoing basis with the limited resources that the Ministry has at its disposal. Other benefits to an "out-of-the-box" solution include:

- large pool of skills available to select from (i.e. software vendor)
- limited programming expertise required
- with version upgrades, solution systems can easily implement new functionality
- puts pressure on software vendors to enhance "out-of-the-box" functionality (as opposed to custom coding)
- develops common skill sets and solution methodologies across an organization.

In January 2003, the *CLAIMaps III* website was launched by the Mineral Development and Lands Branch. This application utilizes ESRI ArcSDE and ArcIMS technologies and ORACLE 8i RDBMS together with Macromedia JRun 3.1 to create maps that are served through Microsoft Internet Information Server on a Mi-

crosoft Windows 2000 Server. The *CLAIMaps* application website, which provides 3 million map images and 70,000 final maps per year, has garnered international recognition and awards by delivering daily updated land tenure maps-to-scale for Ontario.

The graphical user interface of the *CLAIMaps* website is intuitive and easy to use, which are key selling points with MNDM's client base. In addition, clients can link into the ArcIMS map services and incorporate land tenure data into their private, proprietary database systems to generate mining land tenure maps and other complex analytical products in the privacy of their work or home environments.

The *CLAIMaps* website is updated every night through the Automated CLAIMap Management System (ACMS) Replication manager software synchronizes land tenure databases between the internal production server (ACMS) and the public facing internet server (CLAIMaps) thus fulfilling the additional function of providing offsite storage of the Ministry's data. If required, these data can readily be used for business resumption in the event of catastrophic failure of either the ACMS or *CLAIMaps* servers.

The Ministry's key clients have clearly indicated that they would prefer the delivery of the other Ministry products, specifically geoscience data and reports, through an application that operates in a fashion comparable to the *CLAIMaps* website.

Out-of-the-Box Software

The new ***Geology Ontario*** portal website was built using the latest versions of the same software utilized by the *CLAIMaps* website, namely, Oracle 9i, ArcSDE 9, ArcIMS 9, XML, java, JavaScript, and Apache Tomcat (J2EE). Data is served through the Internet Information Server component available in Microsoft Server 2003. The Oracle and ArcSDE software will reside on the database server, while ArcIMS and Windows Internet Information Server will reside on a separate application server.

The *CLAIMaps* website, which currently resides on a Windows 2000 server, will be moved to the *Geology Ontario* servers. All environment variables will be modified so that the *CLAIMaps* application points to the new Oracle/ArcSDE server. The ArcSDE component of the Oracle database will be tuned to optimize performance. The data that reside within the current Oracle 8i databases will be migrated to Oracle 9i. The replication process that currently updates the *CLAIMaps* server will be revised to point to the new Oracle/ArcSDE server so that data can be updated in a similar fashion. The critical components of the new *Geology Ontario* website include:

- Land Information Ontario
- Ontario Land Information Directory
- Internet Geospatial Data Delivery
- XML, ArcXML, Java, Javascript
- ESRI ArcIMS 9, AcSDE 9, ArcGIS 9
- ESRI ArcMap Server, ArcGIS Server, Portal Toolkit
- ESRI Metadata Server, Java ADF
- Apache Tomcat
- Oracle 8i migration to 9i
- Ontario Mining Land Tenure Data
- Geoscience Digital Data
- *CLAIMaps III* Functionality
- ArcIMS
- Windows 2003 Server
- iSERV Production Facilities and the processes.

Highlights

The guiding principle driving this project was to enhance discovery and download of all MNDM geoscience data, maps, reports, and publications. In excess of 300 gigabytes of data have been made available on the *Geology Ontario* web portal. All existing assessment files and publications (over 2 million pages and 168,000 maps), stored as image files, were converted to "PDF searchable" format for more rapid and convenient downloading. As mentioned, the "PDF searchable" format allows for powerful data mining capabilities, since all PDF documents are indexed and can be queried for virtually any text string or combination of text strings. Also, the "PDF searchable" format makes it possible to incorporate government data into a client's work environment using copy/paste functionality (Figure 4). Once downloaded onto a client's personal computer, these documents can be indexed and their contents discoverable using desktop search tools like Google Desktop and Copernic.

Data discovery is based on the ESRI Portal Toolkit and metadata tools (Figure 5). Clients enter search criteria, generate a search, and receive the results in a tabulated format with a description of the documents and their file sizes as well as a hotlink for direct download (Figure 6). In addition, *Geology Ontario* provides links to other Ministry web sites (Figure 7) and map services, which allows clients simultaneous access to services of other government agencies, including those in other Provinces (Figure 8).

For digital data products, such as CAD drawings, ArcView shapefiles, and database files, clients will be required to accept a disclaimer prior to download. The disclaimer is in the form of an html/asp document, controlled by the Internet Information Server component of Windows 2003 server. The Government of Ontario is limited with respect to placing cookies on client computers and, as a result, the disclaimer acceptance avoids leaving cookies.

The Ministry has also developed a map browsing site, much like the *CLAIMaps* website, where clients can simultaneously view numerous thematic datasets (Figure 9). Clients have clearly indicated their preference for viewing

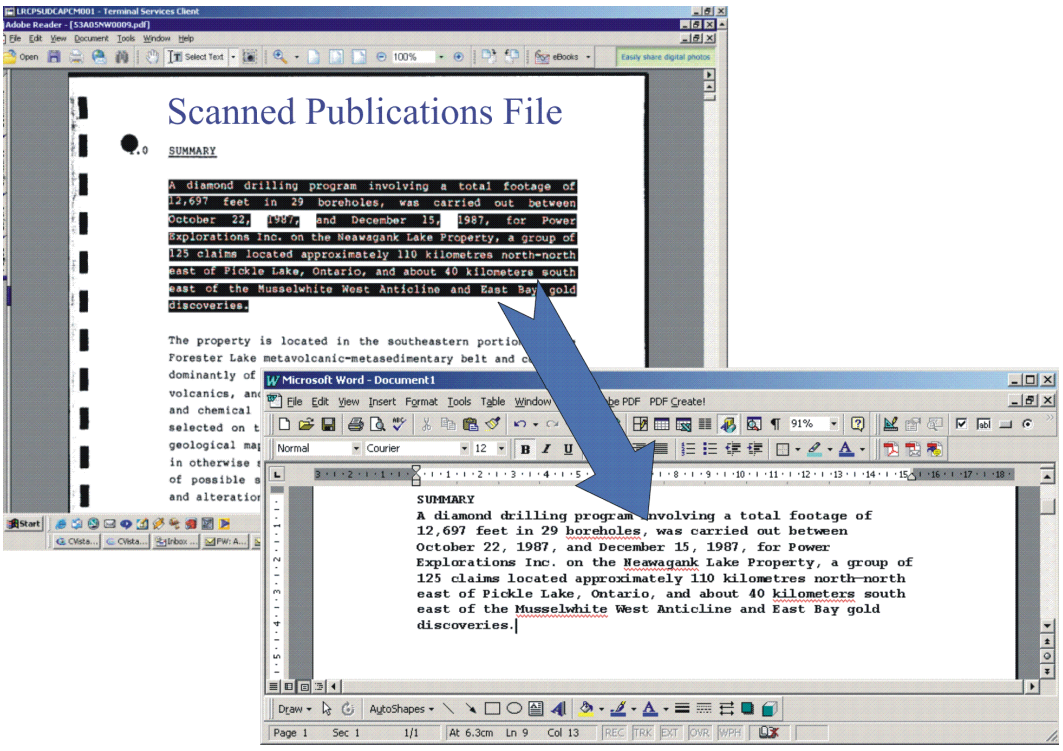


Figure 4. ***Geology Ontario*** "PDF searchable" files allow for copy and paste functionality with all its downloadable documents.

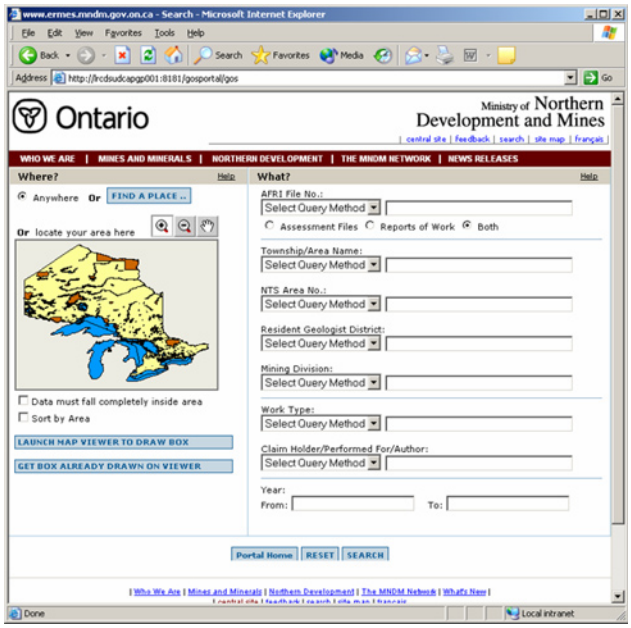


Figure 5. ***Geology Ontario*** text search window for querying the mineral exploration assessment files database (AFRI).

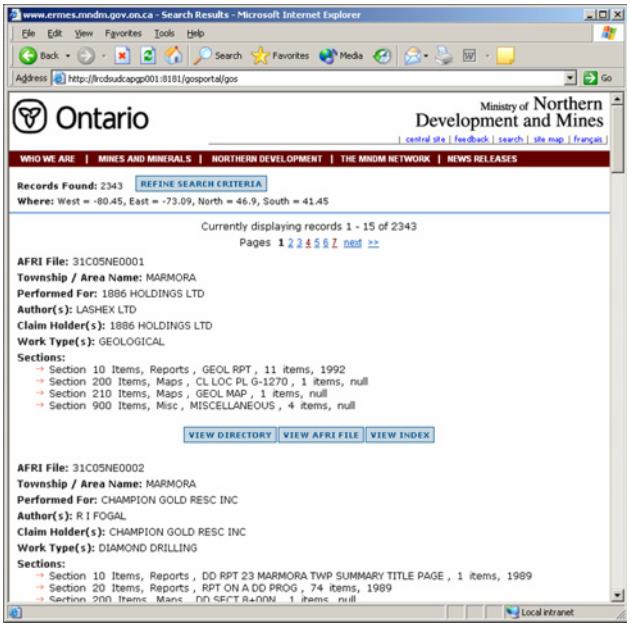


Figure 6. Search results from a typical query of the AFRI database with linkages to detailed metadata and to the folder containing the downloadable files.

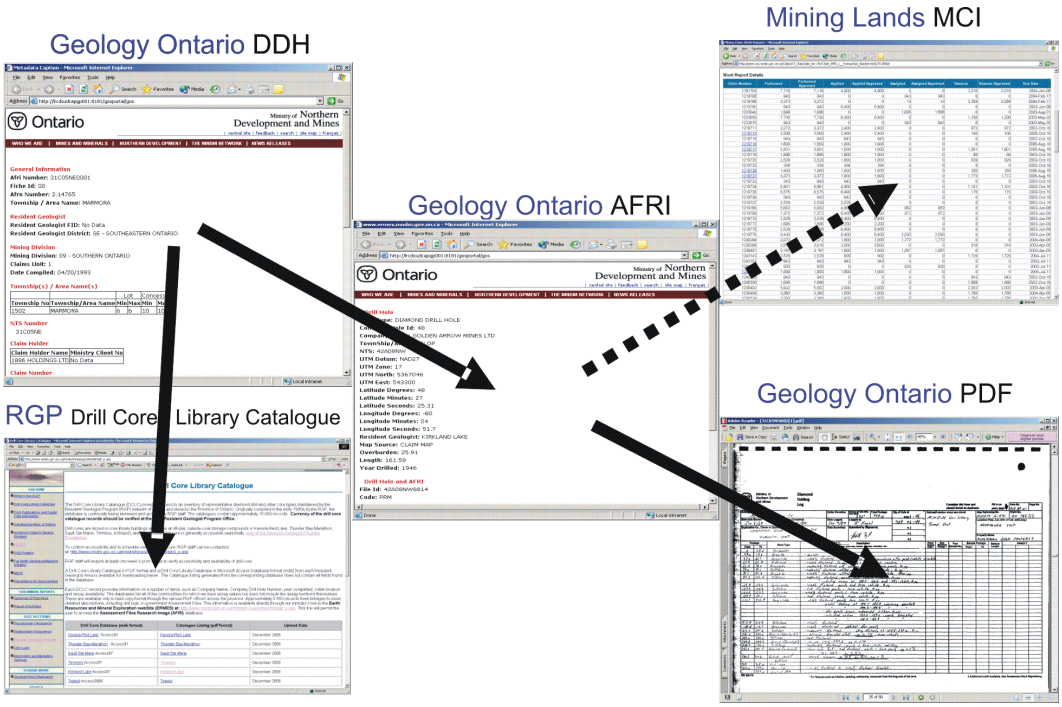


Figure 7. The ***Geology Ontario*** web site provides links to other Ministry web sites – either directly using hypertext (solid arrows) or indirectly using reference numbers in the metadata (dashed arrow). (abbreviations: DDH – diamond drill hole database; RGP – Resident Geologist Program; AFRI – Assessment File Research Imaging database; MCI – Mining Claims Information database; PDF – Portable Document Format).

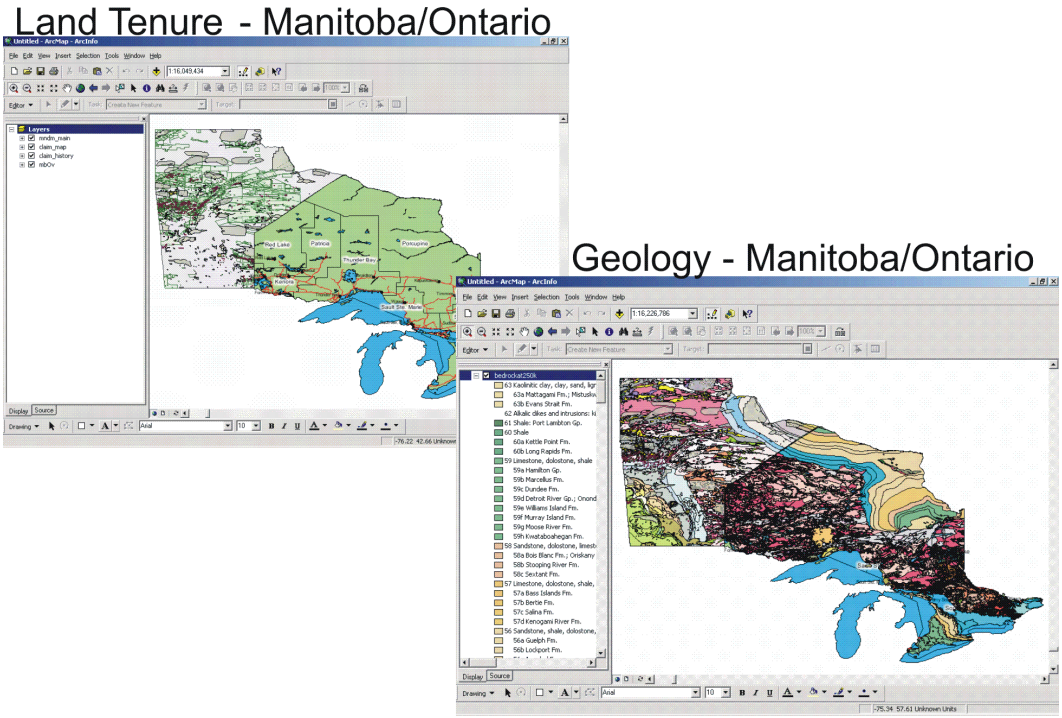


Figure 8. Because ***Geology Ontario*** uses ArcIMS technology, clients can simultaneously link to ArcIMS-based web sites from other jurisdictions – in this illustration land tenure and geology are being accessed concurrently from both the ***Geology Ontario*** web portal and the Manitoba portal.

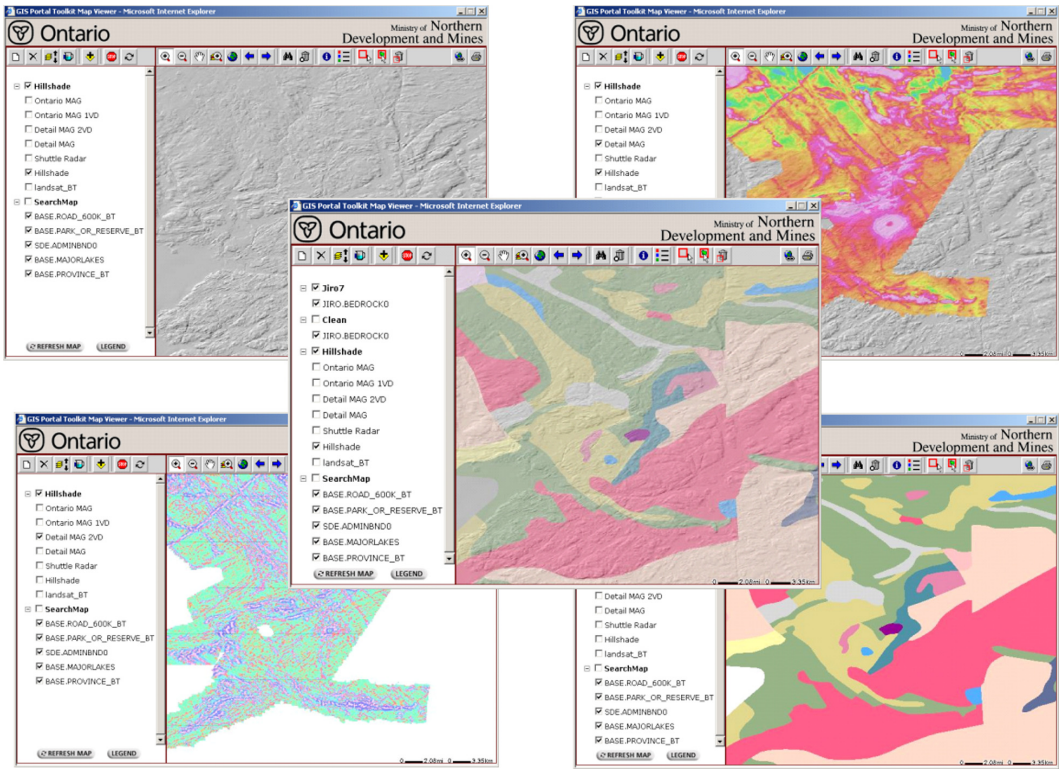


Figure 9. Samples of the ***Geology Ontario*** map browser window illustrating clockwise from top left, a DEM, total field magnetics draped over a DEM, geology, magnetics, and geology draped over a DEM (center image). Over 2 dozen pan-Ontario themes are available for viewing in the map browser window.

mineral land tenure together with geology, geophysics, geochemistry, and other thematic data in order to relate land tenure to known geoscientific characteristics. The GeoPortal Project Team has created geological, geophysical, geochemical, and geomorphologic thematic datasets for multi-theme viewing. Clients can also print maps to scale in a fashion comparable to the *CLAIMaps* website and include various geoscience themes in the map output.

FUTURE WEB DELIVERY

In addition to delivering geoscience data to MNDM clients worldwide, the Ministry is also looking at enhancing the visual web delivery of its data. Some avenues being explored include using 3-dimensional perspective tools with multi-thematic overlay capabilities through a standard web browser (Figure 10) and reviewing best practices in use by other organizations, as far as delivering geoscience data through interfaces such as Google Earth (Figure 11), Microsoft Virtual Earth, NASA World Wind, or ESRI ArcReader and ArcGIS Explorer. Utilizing these additional applications requires careful planning and a balanced client-server approach. Large, static datasets such as geophysical grids digital elevation and remotely sensed imagery can be downloaded and reside on the client side while dynamic datasets such as drilling data and rock sample analyses can be accessed directly from the *Geology Ontario* server through keyhole markup language (kml) or Open Geospatial Consortium (OGC) web mapping services.

ACKNOWLEDGEMENTS

The authors would like to thank Terry Johnston, a member of the *Geology Ontario* project team, for her critical review of this paper.

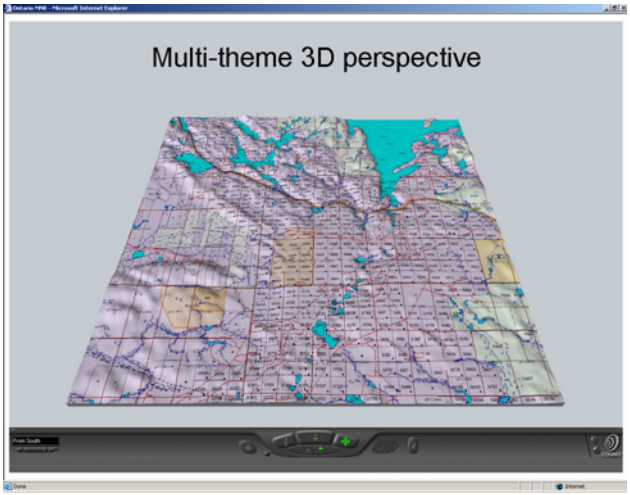


Figure 10. Possible future web viewing enhancements for ***Geology Ontario*** – providing 3-dimensional perspective tools with multi-thematic overlay capabilities through a simple web browser; in this illustration, the mineral land tenure fabric is draped over a DEM.

Figure 11. MNDM is reviewing best practices as far as incorporating ***Geology Ontario*** geoscience data into sites like Google Earth.